NPS-OR-98-004

NAVAL POSTGRADUATE SCHOOL
Monterey, California

A Cascade Approach for Staircase Linear Programs

by

Steven F. Baker
Richard E. Rosenthal

July 1998

Approved for public release; distribution is unlimited.

Prepared for: Air Force Studies & Analyses Agency
1570 Air Force, Pentagon
Washington, DC 20330

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943-5000

RADM Robert C. Chaplin
Superintendent

Richard Elster
Provost

This report was prepared for and funded by the Air Force Studies and Analyses Agency, Washington, DC.

This report was prepared by:

Richard E. Rosenthal

RICHARD E. ROSENTHAL
Professor of Operations Research

REK for

STEVEN F. BAKER, LTCOL, USAF
Assistant Professor, USAF Academy, CO

Reviewed by:

R K Wood for GGB

GERALD G. BROWN
Associate Chairman for Research
Department of Operations Research

Richard E. Rosenthal

RICHARD E. ROSENTHAL
Chairman
Department of Operations Research

Released by:

D W Netzer

DAVID W. NETZER
Associate Provost and Dean of Research

REPORT DOCUMENTATION PAGE

Form approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1998	Technical

4. TITLE AND SUBTITLE

A Cascade Approach for Staircase Linear Programs

5. FUNDING

NMIPR0097927194
NMIPR0098927158

6. AUTHOR(S)

Steven F. Baker and Richard E. Rosenthal

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER

NPS-OR-98-004

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Studies & Analyses Agency
1570 Air Force, Pentagon
Washington, DC 20330

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words.)

We develop a method to approximately solve a large staircase linear program that optimizes decisions over multiple time periods. A bound on the approximation error is also developed. The approximation is derived by a proximal cascade, which sequentially considers overlapping subsets of the model's time periods, or other ordinally defined set. In turn, we bound the cascade's deviation from the optimal objective value by a Lagrangian cascade, which uses proximal cascade dual variables to penalize infeasibility. When tested on the NPS/RAND Mobility Optimizer, a large temporal LP developed for the US Air Force, we often observe gaps between the approximation and bound of less than 10 percent, and save as much as 80 percent of the time required to solve the original problem [Baker, 1997]. We also address methods to reduce the gap, including constraint extension of the Lagrangian cascade, as well as a cut generation approach similar to nested decomposition.

14. SUBJECT TERMS

mobility and transport of forces; large-scale optimization; linear programming

15. NUMBER OF PAGES
28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5800

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18

A CASCADE APPROACH FOR STAIRCASE LINEAR PROGRAMS

Steven F. Baker
Department of Management, US Air Force Academy

Richard E. Rosenthal
Operations Research Department, Naval Postgraduate School

June 19, 1998

abstract

We develop a method to approximately solve a large staircase linear program that optimizes decisions over multiple time periods. A bound on the approximation error is also developed. The approximation is derived by a *proximal cascade,* which sequentially considers overlapping subsets of the model's time periods, or other ordinally defined set. In turn, we bound the cascade's deviation from the optimal objective value by a *Lagrangian cascade,* which uses proximal cascade dual variables to penalize infeasibility. When tested on the NPS/RAND Mobility Optimizer, a large temporal LP developed for the US Air Force, we often observe gaps between the approximation and bound of less than 10 percent, and save as much as 80 percent of the time required to solve the original problem [Baker, 1997]. We also address methods to reduce the gap, including constraint extension of the Lagrangian cascade, as well as a cut generation approach similar to nested decomposition.

1 Introduction

This paper presents a combined method to approximate the solution of a staircase linear program (SLP) and to provide a bound on the approximation's error. The approximation method, or *proximal cascade*, uses a rolling-horizon technique to sequentially solve overlapping subsets of an SLP, where each subset is defined by a contiguous portion of the staircase. The error bound is produced by the *Lagrangian cascade*, which solves subproblems that are also defined by contiguous portions of an SLP, but are made separable by relaxing rows that would otherwise link columns from different subproblems. We provide results of the cascades used on the NPS/RAND Mobility Optimizer (NRMO), a large linear program (LP) that analyzes US Air Force (USAF) airlift effectiveness in a wartime contingency [Rosenthal, *et al.*, 1997].

Staircase linear optimization models are widely used in many areas such as scheduling, where decisions of a given time period directly affect only the decisions of proximal time

periods. Unfortunately, SLPs frequently require considering a large, if not infinite number of time periods. This presents two difficulties: 1) data gathering for the latter periods of such a model may prove problematic, and 2) the resulting model may be too large to solve. Not surprisingly, a human scheduler faces the same difficulties, namely reconciling the increasing number of options with decreasing certainty as the number of time periods grows. For either the human scheduler or the optimization model, perhaps the most straightforward way of dealing with the difficulties incurred by a large problem is to make decisions involving only a subset of the problem's time periods, and then move forward to a new subset. This temporally proximate *myopia*, or inability to see the full future problem at any one point, may result in a suboptimal solution, but can make the problem simple enough to solve. Moreover, a model like NRMO, which attempts to mimic scheduling in order to produce plans but not actual schedules, is better if it can incorporate the realism of myopic scheduling. For example, when NRMO is used to help select aircraft fleets or airbase infrastructure so as to maximize a delivery system's effectiveness, it should optimize based on the inherent limitations imposed by wartime uncertainty, not on a omniscient scheduling capability. Thus, myopia is desirable whenever perfect foresight is unwarranted.

Charnes and Cooper [1961, pp. 370-388] first suggested limiting the number of time periods considered in an SLP, and there have been many variations since. Most notably, Brown, Graves, and Ronen [1987] develop *solution cascading*. This method solves a series of small problems, each consisting of a subset of a large model's time periods. These small problems then form an advanced basis from which the large problem, or monolith, is easily solved. Other, "rolling horizon" applications truncate an infinite time horizon, which consider temporal myopia an unfortunate, but necessary by-product of the truncation. In contrast, the *proximal cascade* proposed here solves overlapping subproblems of models with a finite number of time periods, and those where temporal myopia must be part of the modeling abstraction (Brown, Dell, and Wood [1997] present a number of models where myopia is applicable).

The closeness of a proximal cascade approximation to the overall LP solution is dependent on many scenario-specific factors, and cannot be guaranteed for most problems (significant exceptions to this include Manne [1970], Aronson *et al.* [1985], and Walker [1995]). In order to supplement the proximal cascade approximation, we also develop an optimistic bound on the LP's solution value by exploiting information derived from the proximal cascade. By relaxing the constraints associated with certain time periods of an SLP, we can de-couple a large problem into several subproblems.

Lagrangian relaxation has long been used for decompositions of many sorts; it discourages violation of relaxed constraints through penalties. The Lagrangian penalty is applied

to a series of separable subproblems, and an optimistic bound for the monolith solution's objective value is derived [*e.g.*, Parker and Rardin, 1988, pp. 205-237]. Unfortunately, finding the correct penalty values for relaxed constraints is often as difficult as solving the problem without the relaxation. However, we show that reasonable penalties for the relaxed constraints are readily available from prior dual solutions during the proximal cascade. A *Lagrangian cascade* produces a bound on the LP solution by incorporating the proximal cascade penalties in its subproblems. When combined with the proximal cascade approximation, the size of the gap between the two values gives a quantitative assessment of proximal cascade's solution value.

We outline the notation used in this paper in Section 2. In Sections 3 and 4 we describe the proximal and Lagrangian cascades. In Section 5 we propose a variation on Benders' decomposition relevant to cascades. Finally, in Section 6 we present a NRMO case study that uses cascades.

2 Notation

Except as noted, we assume the following staircase structure for the problem to be solved (referred to as the "monolith," or "problem M"):

$$
\begin{array}{llrrcll r}
(M) & z^* = & \min & \sum_{t\in T} c_t x_t & & & & (1)\\
& & \text{s.t.} & B_t x_{t-1} \;\; +A_t x_t & \geq & b_t & \forall\, t \in T\backslash\{1\} \quad (\alpha_t) & (2)\\
& & & x_t & \geq & 0 & \forall\, t \in T & (3)
\end{array}
$$

where $t \in T$ denotes time periods (or other ordinally defined set), c_t, b_t, A_t, B_t are given data, and x_t, α_t are vectors of primal and dual decision variables. We assume this problem has primal solution x_t^*. Rows and columns that intersect in a non-zero element of A_t or B_t are said to be *associated.* Each row (indexed by t) is associated with columns indexed by t and $t-1$; this creates a *linkage* between these two set elements. A row's linkage of t and $t-1$ denotes a row *width* of two.

Problem M is a simplification of a more general monolithic form, which may include rows with width greater than two. All the results and techniques of this paper have been applied to the more general form, but for simplicity of presentation we retain the row width of two.

3 Proximal Cascade

A proximal cascade is composed of subproblems $1, 2, ..., N$, each of which consists of all rows and columns indexed by overlapping subsets of T, the *cascade index set* (conventionally

a set of time periods). These are the *active* rows and columns of a subproblem n; the corresponding elements of T are the *active indices*. Figure 1 depicts N overlapping subsets of T, suggesting a cascade.

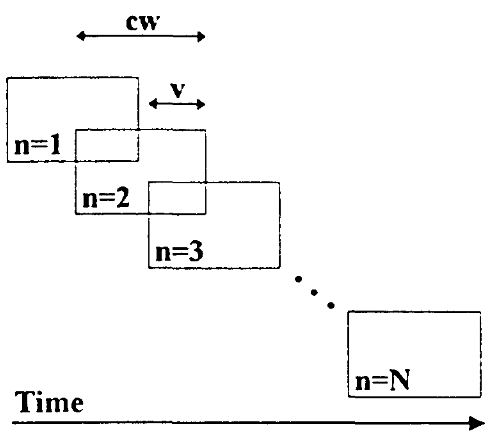

Figure 1: Sequence of subproblems n forming a proximal cascade. Each subproblem contains rows and columns indexed by overlapping subsets of active time periods. The input parameters cw and v specify the number and overlap of active time periods, respectively.

Two input parameters are required to define the cascade subproblems: 1) *cascade width* cw, the number of periods active in each subproblem, and 2) *cascade overlap* v, the number of periods common to two adjacent subproblems. The following additional parameters help describe the subproblems:

ft^n	the first time period of subproblem n.
lt^n	the last time period of subproblem n.
NC	$n \in \{1, ..., N\}$, the set of proximal cascade subproblems.
TC^n	$\{t \in T : ft^n \leq t \leq lt^n\}$, the active periods of subproblem n.
TF^n	$\{t \in T : ft^n \leq t < ft^{n+1}\}$ for $n < N$ $\{t \in T : ft^n \leq t\}$ for $n = N$ the active periods of subproblem n excluding periods that overlap with the next subproblem
x_t^n	solution vector for subproblem n. A feasible solution is assumed to exist.

Throughout the paper, we use a fixed cw and v, so that ft^n and lt^n are derived according to:

$$
\begin{aligned}
ft^n &= (n-1) \cdot (cw - v) + 1 \\
lt^n &= \min\left[T, (n-1) \cdot (cw - v) + cw\right]
\end{aligned}
$$

Each subproblem n has the form:

$$
\begin{array}{llrcllr}
(PCAS^n) & & & & & & \\
z^n = & \sum_{n'<n} \sum_{t \in TF^{n'}} c_t x_t^{n'} + \min \sum_{t \in TC^n} c_t x_t & & & & & (P^n.1) \\
\text{s.t.} & & A_t x_t & \geq & b_t - B_t x_{t-1}^{n-1} & t = ft^n & (P^n.2) \\
 & B_t x_{t-1} & +A_t x_t & \geq & b_t & ft^n < t \leq lt^n & (P^n.3) \\
 & & x_t & \geq & 0 & ft^n \leq t \leq lt^n & (P^n.4)
\end{array}
$$

In addition to the active variables, the objective function of a $PCAS^n$ subproblem includes variables indexed by $t < ft^n$, which are *fixed* to the optimal value computed in the last subproblem in which they were active. The first constraint's right-hand-side is reduced by the resources consumed by the fixed level of x_{t-1}^{n-1}. Thus, the feasibility of $PCAS^n$ depends on x_{t-1}^{n-1}. The constraints indexed by $t : ft^n < t \leq lt^n$ are unchanged, and the constraints indexed by $t > lt^n$ are relaxed. Figure 2 shows the relationship of a proximal cascade subproblem to the surrounding subproblems.

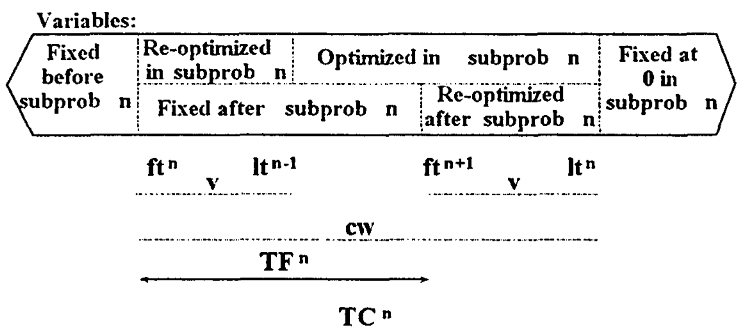

Figure 2: A single proximal cascade subproblem optimizes variables whose time periods are indexed by the active set ($t \in TC^n$). Thus, it re-optimizes variables indexed by time periods active in the previous subproblem, $t \in TC^{n-1} \cap TC^n$ (the number of periods in this overlap is v). Constraints of future time periods are relaxed, and variables of future periods are fixed at level 0. Variables whose time periods are indexed by $t \in TF^n$ are fixed after subproblem n.

The proximal cascade heuristic proceeds as follows :

For $n = 1, ..., N$ {
  Define and solve subproblem $PCAS^n$
  Fix and output the value of $x_t^n, \forall t \in TF^n$
}
Output proximal cascade objective function value, z^N.

Selection of cascade width, cw, and cascade overlap, v can play a large role in the cascade solution quality, which we define as $1/\left|z^N - z^*\right|$ (solution quality is infinite when $z^N = z^*$). Our experimental results suggest that v should be at least as large as the maximum number of time period indices that are common to consecutively indexed rows. This permits every row to have all associated columns active in at least one subproblem.

The following proposition demonstrates that the proximal cascade's objective function value provides an upper bound on the monolith's objective function value:

Proposition 1 $\quad z^* \leq z^N$.

Proof:

$$\begin{aligned}
z^* = \quad & \min && \sum_{t \in T} c_t x_t \\
& \text{s.t.} && (2), (3) \\
\leq \quad & \min && \sum_{t \in T} c_t x_t \\
& \text{s.t.} && (2), (3) \\
& && x_t = x_t^n \quad \forall n < N, t \in TF^n \\
= \quad & && \sum_{n' < N} \sum_{t \in TF^{n'}} c_t x_t^{n'} + \min \sum_{t \in TC^N} c_t x_t \\
& \text{s.t.} && (P^N.2), (P^N.3), (P^N.4), \\
= \quad & z^N. &&
\end{aligned}$$

The inequality holds because fixing a subset of the x_t restricts the original problem.

□

In addition to providing an upper bound on z^*, a feasible proximal cascade solution $(x_t^n, \forall\, n \in NC,\ t \in TF^n)$ is feasible to the monolith, since the rows of the monolith are enforced by the rows of $PCAS^n\ \forall\, n \in NC$.

The primary advantages of using a proximal cascade lie in its abilities: 1) to approximate the solution of arbitrarily large staircase LPs, and 2) to mimic real-world myopia when appropriate. However, since the proximal cascade is a heuristic, some bound on the solution error is desirable.

4 Lagrangian Cascade

Lagrangian relaxation has long been used to bound, or even solve, perhaps approximately, linear and integer programs by solving subproblems. A temporal partition allows each subproblem to be solved separately by taking the Lagrangian relaxation of rows that would otherwise appear in more than one subproblem. The structure of a staircase problem invites

Lagrangian relaxation, since rows are associated with variables from only a small number of adjacent time periods. It is all the more attractive if the maximum row width is small.

Like a proximal cascade, a Lagrangian cascade is composed of subproblems, each consisting of active rows and columns indexed by subsets of the cascade index set T. Unlike the proximal cascade, the active index subsets are disjoint, forming a partition of T. The monolith problem is separated into Lagrangian subproblems by relaxing rows that link columns with cascade indices of different subsets. As with all Lagrangian relaxations, objective function penalties are added to the columns of these relaxed constraints in order to discourage violation. We refer to these rows as *Lagrange-relaxed* rows.

Unlike traditional Lagrangian relaxation, a Lagrangian cascade may exploit a previously computed proximal cascade's dual solution for its objective function penalties. This obviates a computationally intensive multiplier search.

Figure 3 illustrates a partition of T into L subsets, each of which has width *lwid,* except for the last subset which may be limited by the cardinality of T . Note that *lwid* need not be constant for all subproblems, but is presented that way for convenience.

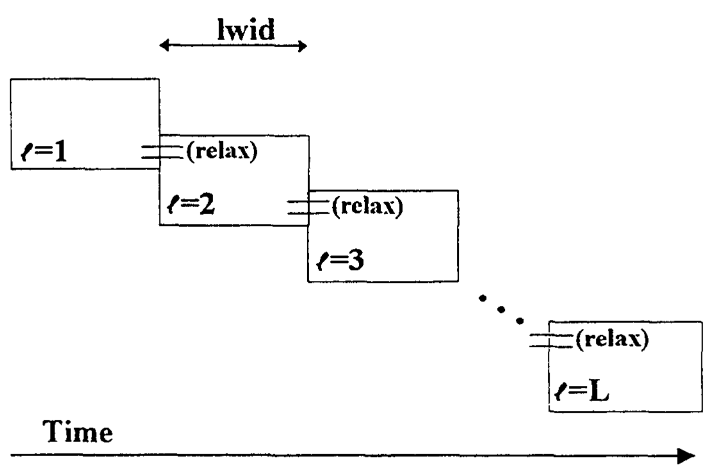

Figure 3: A Lagrangian cascade partitions the rows and columns of a monolith into many Lagrangian subproblems ℓ of contiguous time periods. Overlapping rows are Lagrange-relaxed.

The following additional notation is used to describe the Lagrangian cascade:

ft^ℓ	The first time period in subproblem ℓ.
lt^ℓ	The last time period in subproblem ℓ.
TRL^ℓ	$\{t : ft^\ell \leq t \leq lt^\ell\}$, the active periods of subproblem ℓ
TR	$\left\{ft^2, ft^3, ..., ft^L\right\}$ the periods corresponding to Lagrange-relaxed row indices
TL^ℓ	$\begin{cases} TRL^\ell, \ell = 1 \\ \{t : ft^\ell < t \leq lt^\ell\},\ \ell > 1 \end{cases}$ the periods of subproblem ℓ corresponding to active row indices
CL	$\ell \in \{1, ..., L\}$, the set of Lagrangian cascade subproblems
$I(\cdot)$	1 if argument is true; 0 otherwise

As before, specifying the width value *lwid* permits a simple derivation of ft^ℓ and lt^ℓ :

$$ft^\ell = (\ell - 1) \cdot lwid + 1$$

$$lt^\ell = \min\left[T, \ell \cdot lwid\right]$$

Figure 4 shows the relationship of a Lagrangian cascade subproblem to its neighboring subproblems.

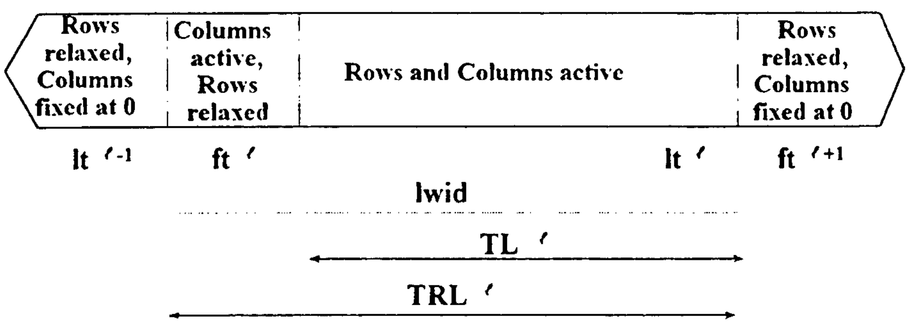

Figure 4: A single Lagrangian cascade subproblem includes columns indexed by $t \in TRL^\ell$ (the active set), and rows indexed by $t \in TL^\ell$. Rows indexed by ft^ℓ are relaxed, and a Lagrangian penalty is applied to the objective function coefficients of active columns associated with relaxed rows.

Given this notation, define the Lagrangian cascade problem LC:

$$
\begin{array}{llllll}
(LC) & & & & & \\
z^{LC} = & \min & \sum_{t \in T} c_t x_t & & & \\
 & & + \sum_{t \in TR} \alpha_t \left(b_t - A_t x_t - B_t x_{t-1}\right) & & & (LC.1) \\
 & \text{s.t.} & B_t x_{t-1} \quad + A_t x_t & \geq b_t & \forall t \in \cup_\ell TL^\ell & (LC.2) \\
 & & \qquad\qquad x_t & \geq 0 & \forall t \in T & (LC.3)
\end{array}
$$

The objective function includes Lagrangian penalties $\alpha_t \geq 0$. The remaining structural constraints include only the active staircase rows.

Because all of the linking rows between subproblems are Lagrange-relaxed, LC decomposes into L disjoint subproblems with $z^{LC} = \sum_\ell z^\ell$:

$$
\begin{array}{llllll}
(LCAS^\ell) & & & & & \\
z^\ell = & \min & \sum_{t \in TRL^\ell} c_t x_t + \alpha_{ft^\ell}\left(b_{ft^\ell} - A_{ft^\ell} x_{ft^\ell}\right) \cdot I(\ell > 1) & & & \\
& & -\alpha_{ft^{\ell+1}} B_{ft^{\ell+1}} x_{lt^\ell} \cdot I(\ell < L) & & & (L^\ell.1) \\
& \text{s.t.} & B_t x_{t-1} \quad +A_t x_t \ \geq \ b_t & & \forall t \in TL^\ell & (L^\ell.2) \\
& & \qquad\qquad\quad x_t \ \geq \ 0 & & \forall t \in TRL^\ell & (L^\ell.3)
\end{array}
$$

A Lagrangian cascade proceeds as follows:

Record dual variable levels (α_t) from proximal cascade solution $\forall t \in TR$.
For $\ell = 1, 2, ..., L$ {
  Define and solve subproblem $LCAS^\ell$ given above
  Record the value of z^ℓ
}
Output the Lagrangian cascade solution value: $\sum_\ell z^\ell$.

Solving the relaxed problem in this manner allows the tractable computation of a lower bound on z^*. By the theorem of weak Lagrangian duality [*e.g.*, Parker and Rardin, 1988, p. 206],

$$z^{LC} = \sum_\ell z^\ell \leq z^*.$$

As stated earlier, the quality of this bound depends in large measure on the quality of the dual variables. These variables, in turn, depend on the quality of the proximal cascade solution. Frequently, the closer a proximal cascade solution approximates the optimal monolith solution, the closer the associated duals will approximate the optimal monolith dual solution. Hence, there is strong incentive for making the proximal cascade solution as close to the monolith solution as possible.

4.1 Extended Constraints

The Lagrangian cascade bound can be improved by modifying the constraint: $B_{ft^\ell} x_{lt^{\ell-1}} + A_{ft^\ell} x_{ft^\ell} \geq b_{ft^\ell}$, which is Lagrange-relaxed in each subproblem ℓ. However, to insure the validity of the relaxation, we define a surrogate copy of the variable $x_{lt^{\ell-1}}$, which we denote $\tilde{x}_{lt^{\ell-1}}$.

Consider problem $\widetilde{M}$, which is identical to the monolith, but with constraint $\widetilde{M}.1$ and $\widetilde{M}.2$ added.

$$
\begin{array}{llllllll}
(\widetilde{M}) & z^{\widetilde{M}} = & \min & \sum_{t\in T} c_t x_t & & & & (1) \\
& & \text{s.t.} & B_t x_{t-1} & +A_t x_t & \geq & b_t \quad 1 < t \leq |T| & (2) \\
& & & & x_t & \geq & 0 \quad \forall t \in T & (3) \\
& & & B_t \tilde{x}_{t-1} & +A_t x_t & \geq & b_t \quad \forall t \in TR & (\widetilde{M}.1) \\
& & & \tilde{x}_{t-1} & & \geq & 0 \quad \forall t \in TR & (\widetilde{M}.2)
\end{array}
$$

$\widetilde{M}.1$ restores all Lagrange-relaxed constraints from LC, but it replaces x_{t-1} with a surrogate variable $\tilde{x}_{t-1}$ $\forall t \in TR$. Except for non-negativity, $\tilde{x}_{t-1}$ appears nowhere else in the formulation. By the following proposition, $\widetilde{M}$ is not a restriction of the monolith.

Proposition 2 $\qquad z^* \geq z^{\widetilde{M}}$

Proof: Let $\tilde{x}_{t-1} = x^*_{t-1}$ $\forall t \in TR$. Since x^*_{t-1} must be feasible to (2) and (3), it must also be feasible to $\widetilde{M}.1$ and $\widetilde{M}.2$, and consequently problem $\widetilde{M}$. Thus, $z^{\widetilde{M}}$ can be no worse than z^*.

□

In fact, $z^* = z^{\widetilde{M}}$, because the surrogate columns do not contribute to the objective, nor do they allow the original columns to further contribute to the objective. However, this is not central to the overall result, which is to show that a Lagrangian relaxation of $\widetilde{M}$ is still a relaxation of the monolith M. We define this relaxation as $\widetilde{LC}$ (problem $\widetilde{M}$ with constraint (2) relaxed for all $t \in TR$):

$$
\begin{array}{llllllll}
(\widetilde{LC}) & & & & & & & \\
z^{\widetilde{LC}} = & \min & \sum_{t\in T} c_t x_t & & & & & \\
& & + \sum_{t\in TR} \alpha_t (b_t - A_t x_t - B_t x_{t-1}) & & & & & \\
& \text{s.t.} & B_t x_{t-1} & +A_t x_t & \geq & b_t & \forall t \in \cup_\ell TL^\ell & (LC.2) \\
& & & x_t & \geq & 0 & \forall t \in T & (LC.3) \\
& & B_t \tilde{x}_{t-1} & +A_t x_t & \geq & b_t & \forall t \in TR & (\widetilde{M}.1) \\
& & \tilde{x}_{t-1} & & \geq & 0 & \forall t \in TR & (\widetilde{M}.2)
\end{array}
$$

By combining Proposition 2 with the fact that $\widetilde{LC}$ is a relaxation of $\widetilde{M}$, we have:

$$z^* \geq z^{\widetilde{M}} \geq z^{\widetilde{LC}}.$$

The following proposition shows that $z^{\widetilde{LC}}$ bounds z^{LC} from above, which may improve the Lagrangian bound on z^*:

Proposition 3 $z^{\widetilde{LC}} \geq z^{LC}$

Proof: Relaxing $\widetilde{M}.1$ and $\widetilde{M}.2$ eliminate $\widetilde{x}_t$ from the problem. What remains is problem LC. Thus, LC is a relaxation of $\widetilde{LC}$, and $z^{\widetilde{LC}} \geq z^{LC}$.

□

Combining the above results, we see that $z^{\widetilde{LC}}$,may provide a tighter bound on z^* than z^{LC}:

$$z^* \geq z^{\widetilde{LC}} \geq z^{LC}.$$

Although incorporating extended constraints enlarges each subproblem, the increased solution time may be rewarded by a tighter bound on z^*. Consider the following staircase LP (shown as a maximization):

$$\begin{array}{rlrrrcll}
z^* = & \max & 2X_1 & +4X_2 & +X_3 & & & \\
& \text{s.t.} & X_1 & & & \leq & 2 & \\
& & X_1 & +X_2 & & \leq & 3 & (\alpha_2) \\
& & & X_2 & +X_3 & \leq & 4 & \\
& & X_1, & X_2, & X_3 & \geq & 0. &
\end{array}$$

A solution to this problem is: $X_2^* = 3$, $X_3^* = 1$, with $z^* = 13$. Lagrangian relaxation of the second row results in the following for $\alpha_2 \geq 0$:

$$\begin{array}{rlrrrcl}
z^{LC} = & \max & 2X_1 & +4X_2 & +X_3 & + & \alpha_2(3 - X_1 - X_2) \\
& \text{s.t.} & X_1 & & & \leq & 2 \\
& & & X_2 & +X_3 & \leq & 4 \\
& & X_1, & X_2, & X_3 & \geq & 0.
\end{array}$$

When $\alpha_2 = 1$, the above may be rewritten as

$$\begin{array}{rlrcl}
z^{LC} = 3 & +\max & X_1 & & \\
& \text{s.t.} & X_1 & \leq & 2 \\
& & X_1 & \geq & 0
\end{array}
\qquad
\begin{array}{lrrcl}
+\max & 3X_2 & +X_3 & & \\
s.t. & X_2 & +X_3 & \leq & 4 \\
& X_2, & X_3 & \geq & 0.
\end{array}$$

This has a solution $\overline{X}_1 = 2$, $\overline{X}_2 = 4$, with $z^{LC} = 17$, which is an optimistic bound on the first problem, $z^* = 13$. However, the bound may be tightened by duplicating X_1 with $\widetilde{X}_1$, and incorporating the method of extended constraints:

$$\begin{array}{rlrcl}
z^{\widetilde{LC}} = 3 & +\max & X_1 & & \\
& \text{s.t.} & X_1 & \leq & 2 \\
& & X_1 & \geq & 0
\end{array}
\qquad
\begin{array}{lrrrcl}
+\max & & 3X_2 & +X_3 & & \\
s.t. & \widetilde{X}_1 & +X_2 & & \leq & 3 \\
& & X_2 & +X_3 & \leq & 4 \\
& \widetilde{X}_1 & X_2, & X_3 & \geq & 0.
\end{array}$$

This has solution $X_1 = 2$, $X_2 = 3$, with $z^{\widetilde{LC}} = 14$, resulting in a tighter bound than z^{LC}.

5 A Nested-Cascade Variation of Benders' Decomposition

In this section, we consider the use of successive proximal cascades, which we define as a proximal cascade *series*. We show that cascade solution quality can improve significantly using a strategy that adds weak cuts from previous proximal cascade solutions. This approach is a heuristic variation of Benders' decomposition [Benders 1962].

Exploiting a staircase structure to decompose an LP is described by Glassey [1973], as well as Ho and Manne [1974]. These methods successively add dual cuts to a series of master problems. Each master problem serves as a cut-generating subproblem for another master, thus warranting the name "nested decomposition." Using a variation of this approach in concert with Lagrange multipliers taken from previous proximal cascade dual solutions, we attain a tighter gap between the proximal and Lagrangian cascade solutions than can be obtained in a single-series cascade.

In order to demonstrate nested-cascade decomposition, consider problem $BCAS^n$, which is the solution to the remaining periods, given the fixed columns of subproblems $1,\ldots,n-1$. In other words, $BCAS^n$ provides the solution to the remaining monolith, given the cascade solution for $t \in \bigcup_{n'<n} TF^{n'}$:

$(BCAS^n)$

$$
\begin{array}{llrcll}
z^n = & \sum\limits_{n'<n} \sum\limits_{t\in TF^{n'}} c_t x_t^{n'} & & & & \\
& + \min\limits_{x} \sum\limits_{t\in TC^n} c_t x_t + f(x_{lt^n}) & & & & \\
\text{s.t.} & A_t x_t & \geq & b_t - B_t x_{t-1}^{n-1} & t = ft^n & (BCAS^n.1) \\
& B_t x_{t-1} \quad + A_t x_t & \geq & b_t & ft^n < t \leq lt^n & (BCAS^n.2) \\
& x_t & \geq & 0 & \forall t \in TC^n & (BCAS^n.3)
\end{array}
$$

where:

$$
\begin{array}{llrcll}
f(x_{lt^n}) = & & \min\limits_{x} \sum\limits_{t>lt^n} c_t x_t & & & \\
\text{s.t.} & & A_t x_t & \geq & b_t - B_t x_{lt^n} & t = lt^n + 1 \\
& B_t x_{t-1} & + A_t x_t & \geq & b_t & \forall t > lt^n + 1 \\
& & x_t & \geq & 0 & \forall t \geq lt^n + 1
\end{array}
$$

We can rewrite $BCAS^n$ by taking the dual of $f(x_{lt^n})$:

$$
z^n = \sum_{n'<n} \sum_{t \in TF^{n'}} c_t x_t^{n'} + \min_x \left[\begin{array}{lrl} \sum_{t \in TC^n} c_t x_t + \max_\alpha \alpha_{lt^n+1} (b_{lt^n+1} - B_{lt^n+1} x_{lt^n}) + \sum_{t > lt^n+1} \alpha_t b_t & & \\ \text{s.t.} \quad (BCAS^n.1),(BCAS^n.2),(BCAS^n.3) & & \\ \alpha_t A_t + \alpha_{t+1} B_{t+1} \leq & c_t & lt^n + 1 \leq t < T \\ \alpha_T A_T \leq & c_T & \\ \alpha_t \geq 0 \; \forall t > lt^n & & \end{array} \right]
$$

This formulation is equivalent to:

$$
\begin{aligned}
z^n = & \sum_{n'<n} \sum_{t \in TF^{n'}} c_t x_t^{n'} + \min_x \sum_{t \in TC^n} c_t x_t \\
& + \max_{1 \leq j \leq |J^n|} \alpha_{lt^n+1}^{(j)} (b_{lt^n+1} - B_{lt^n+1} x_{lt^n}) + \sum_{t > lt^n+1} \alpha_t^{(j)} b_t \\
\text{s.t.} \quad & (BCAS^n.1),(BCAS^n.2),(BCAS^n.3)
\end{aligned}
$$

where $\alpha_t^{(j)}$ is a component of vector $\alpha^{(j)} \in J^n$, defined by the region:

$$
\alpha_t A_t + \alpha_{t+1} B_{t+1} \leq c_t \quad lt^n + 1 \leq t < T
$$

$$
\alpha_T A_T \leq c_T
$$

$$
\alpha_t \geq 0 \; \forall t > lt^n
$$

This region, if not bounded, may require that feasibility cuts be added to the formulation of $BCAS^n$ [*e.g.*, Parker and Rardin, 1988, pp.237-244]. Ignoring this complication for now, cascade $BCAS^n$ may be rewritten as:

$$
z^n = \sum_{n'<n} \sum_{t \in TF^{n'}} c_t x_t^{n'} + \min_{x,\theta} \sum_{t \in TC^n} c_t x_t + \theta
$$

$$
\text{s.t.} \quad (BCAS^n.1),(BCAS^n.2),(BCAS^n.3)
$$

$$
\theta \geq \alpha_{lt^n+1}^{(j)} (b_{lt^n+1} - B_{lt^n+1} x_{lt^n}) + \sum_{t > lt^n+1} \alpha_t^{(j)} b_t \qquad j = 1, ..., |J^n| \qquad (BCAS^n.4)
$$

In a traditional nested decomposition, each of these problems serves as the master problem for its successor and the subproblem for its predecessor. The subproblems derive cuts of the form given by $BCAS^n.4$. Accordingly, a relaxed Benders' master problem consists of a subset of these cuts, which is an approximation of the monolith when $n = 1$. Additionally, the master problem includes feasibility cuts (not shown), which ensure that the value of x_{lt^n} permits feasibility of the successor subproblem.

Nested decomposition partitions the rows and columns of the monolith into subproblems. In this way, $TC^n \cap TC^{n+1} = \emptyset$. It follows that $lt^n + 1 = ft^{n+1}$. This does not hold for a proximal cascade, from which nested-cascade decomposition is derived.

The nested-cascade decomposition uses an overlapping series of subproblems, which weaken the optimality cuts given by *$BCAS^n$.4*, but which reduce or eliminate the need for feasibility cuts. In order to distinguish the difference between nested and nested-cascade decomposition, define the latter subproblems as BC^1, BC^2,..., BC^n, BC^{n+1},..., BC^N. Although each subproblem's form is identical to $BCAS^n$, each subproblem is overlapping such that $ft^{n+1} < lt^n + 1$:

$$
\begin{aligned}
&(BC^{n+1}) \\
&z^n = \sum_{n'<n+1} \sum_{t \in TF^{n'}} c_t x_t^{n'} + + \min_x \sum_{t \in TC^{n+1}} c_t x_t + f(x_{lt^{n+1}}) \\
&\qquad \text{s.t.} \quad (BCAS^{n+1}.1), (BCAS^{n+1}.2), (BCAS^{n+1}.3)
\end{aligned}
$$

Unlike the $BCAS^{n+1}$ subproblem, BC^{n+1} re-optimizes $x_{ft^{n+1}},...,x_{lt^n}$. Taking the dual of BC^{n+1} yields a feasible region defined by $\tilde{J}^n$:

$$\alpha_t A_t + \alpha_{t+1} B_{t+1} \leq c_t \quad ft^{n+1} \leq t < T$$

$$\alpha_T A_T \leq c_T$$

$$\alpha_t \geq 0 \; \forall t \geq ft^{n+1}$$

Since $ft^{n+1} < lt^n + 1$, the feasible region defined by $\tilde{J}^n$ is a restriction of the region defined by J^n. In particular, the variable passed to the master problem, α_{lt^n+1}, is more restricted in $\tilde{J}^n$ than in J^n. This can be seen by noting that α_{lt^n+1} is restricted by $\alpha_{lt^n+1} A_{lt^n+1} + \alpha_{lt^n+2} B_{lt^n+2} \leq c_{lt^n+1}$ in both J^n and $\tilde{J}^n$, but also by $\alpha_{lt^n} A_{lt^n} + \alpha_{lt^n+1} B_{lt^n+1} \leq c_{lt^n}$ in $\tilde{J}^n$. Therefore, the cuts provided by subproblem BC^{n+1} are weak, because a restricted dual feasible region corresponds to a relaxed primal feasible region.

It is only the cascade overlap that distinguishes the regions J^n and $\tilde{J}^n$, which suggests that they share many similarities. The result is a trade-off; overlapping subproblems reduces or eliminates the need for feasibility cuts, at a cost of weaker optimality cuts.

The nested-cascade variation of Benders' decomposition proceeds as follows :

Select acceptable gap tolerance, $GAPTOL$
Select maximum number of allowable series, $SMAX$
Let $j = 0$, $gap = M$, $SET\_J = \emptyset \;\; \forall n$
While $j < SMAX$, and $gap > GAPTOL$ {
  For $n = 1, 2, ..., N$ {
    Solve BC^n and record $\alpha_{ft^n}^{(j+1)}$, fix and record $x_t^n \; \forall t \in TF^n$

}
Record z^N
$j = j + 1$
$SET\_J = SET\_J \cup \{j\}\ \forall n$
For $\ell = 1, 2, ..., L$ {
Solve $LCAS^\ell$ using $\alpha_t^{(j)}\ \forall t \in TR$
Record z^ℓ
}
$gap = z^N - \sum_\ell z^\ell$
}
Output solution $x_t^n\ \forall t \in TF^n, n \in NC$ with objective value z^N

We applied the nested-cascade technique to 10 test staircase problems [Baker, 1997]. The first series of the nested-cascade decomposition solved subproblems $BC^1, BC^2, ..., BC^N$, without any cuts. Subsequent series solved these subproblems in the same order using the heuristic cuts generated by the dual variables from subproblems of all previous series. Each series included one additional cut per subproblem. The proximal cascade solution value was the objective value of the last subproblem of the most recent series. The Lagrangian cascade used the dual variables supplied by the most recent proximal cascade. In general, the method did not converge to monolith optimal, but stabilized to an average proximal-Lagrangian gap of 2.7 percent for the 10 problems. Over half (60%) of the gap reduction was attributable to the Lagrangian cascade, which reflected the benefit of more accurate Lagrangian penalties.

The above results suggest a promising alternative to a single series cascade (although this method should not be used when enforcing myopia). Unlike traditional nested decomposition for staircase models [Glassey, 1973; Ho and Manne, 1974], the nested-cascade variation lacks a convergence proof. However, traditional nested decompositions have no cascade overlap, and will often have greater difficulty maintaining primal feasibility. Thus, nested-cascade decomposition has an advantage over many nested methods, which must rely on feasibility cuts.

6 Case Study: The NPS/RAND Mobility Optimizer

The Naval Postgraduate School / RAND Mobility Optimizer was developed in 1996 as an alternative and complement to simulation for USAF strategic airlift analysis [Rosenthal, *et al.*, 1997; Melody *et al.*, 1997]. It is the consolidation of mobility optimization models from the Naval Postgraduate School [Morton, Rosenthal, and Lim, 1996; Rosenthal, *et al.*, 1997] and RAND [Killingsworth and Melody, 1994]. The project's sponsor is the USAF Studies and Analyses Agency, Global Mobility Branch. NRMO has been used by this agency as

well as by the office of the Secretary of Defense [Stucker and Melody, 1997], and the Joint Chiefs of Staff Force Projection Branch [Damm, forthcoming].

Strategic airlift is defined as: "...the movement of units, personnel and material in support of all Department of Defense agencies between the continental United States and overseas areas" [Dept. of the Air Force, 1992, p. 301]. Although this definition embodies many missions, a primary goal of strategic airlift is to maximize the on-time delivery of combat and support forces as directed by the national command authorities. NRMO represents strategic airlift as a multi-period, multi-commodity, network-based LP with a large number of side constraints. The model is used by defense planners to provide insight into mobility issues such as the adequacy of aircraft fleet and airbase infrastructure, as well as the identification of system bottlenecks. Multiple scenarios have been used to address questions of fleet selection and airfield improvements.

There are four primary input requirements of the NRMO LP: 1) the required cargo and passenger movements as delineated by the Time Phased Force Deployment Document (TPFDD), a widely used planning database, 2) the types and numbers of available aircraft and crews, 3) the usable airfields, and 4) the allowable routes for each aircraft type. The LP minimizes the weighted sum of late and undelivered cargo penalties, subject to restrictions such as aircraft flow balance, aircraft payload, and airfield capacity. The solution specifies the airlift mission assignments by requirement moved, aircraft and route flown, and time delivered. From this output, information such as unit closure (the time when all of a unit's cargo and passengers have been delivered) may be computed. Return routings and airfield saturation levels are also given in the LP solution, as well as the marginal values of resources.

In addition to the four primary inputs, other data allow NRMO to model aerial refueling, geographic crew movement, and intra-theater airlift. If directed by the scenario input, NRMO can assign dual-role aircraft as either airlifters or aerial refueling tankers, and reassign them as the contingency warrants. The movement of crews can be modelled geographically by balancing their flow through selected rest bases, and observing limits on crew utilization. Finally, NRMO allows intra-theater activity by alternating selected aircraft between tactical (short-haul) and strategic (long-haul) roles, again as the contingency warrants.

The structure and complexity of NRMO motivated the development of the proximal and Lagrangian cascades, and provided a test-bed of problems. A moderately sized scenario involving over two hundred military units requiring movement results in an LP with around 27,000 rows, 126,500 columns, and 921,500 nonzero coefficients. Problem dimensions can increase well beyond this size as the time-step is decreased and the horizon is increased. Large scenarios can easily overwhelm current computing capabilities. Additionally, the

model should produce results that are intentionally myopic, since change and uncertainty are characteristics of the underlying airlift scheduling system.

NRMO is, for the most part, an SLP with typical row widths of three time periods, which is the typical maximum mission duration. Consequently, the proximal cascade overlap, v must be at least two periods. Correspondingly there must be two periods of Lagrangian-relaxed rows between each subproblem. The model also requires other minor modifications to accommodate cascades.

Two NRMO problem instances are used to test cascades. The first problem is the primary test scenario used at the Naval Postgraduate School to verify and validate air mobility linear programs [Baker, 1997]. The other scenario considers an ongoing study by RAND [Stucker and Melody, 1997]. Extended constraints (Section 4.1) were used in both scenarios; the cascading variation of Benders' decomposition (Section 5) was not used in order to preserve myopia.

The performance tests measure the effect of three parameters on the proximal-Lagrangian gap. Typically, larger values of the proximal cascade width, cw, proximal cascade overlap, v, and Lagrangian cascade width, *lwid* should all reduce the gap. The test results reflect these generalizations.

Both of the problem instances are generated by GAMS [Brooke, *et al.,* 1992], and written into MPS format. Additionally, the GAMS output provides a file that maps each row and column to its associated time index. The cascade logic is written in *C* using the *CPLEX* callable library version 3.0 [*CPLEX*, 1994]. A utility translates the solution reported by *CPLEX* to a GAMS compatible format for further processing. Unless otherwise noted, the computer used is an IBM RS6000/590 with 512MB of RAM. All times are given in CPU seconds.

6.1 Notional Southwest Asia Scenario

The notional Southwest Asia (SWA) scenario is a small, easily solved problem that was originally designed to test THRUPUT II [Morton *et al.*, 1994], one of NRMO's predecessors. It includes 21 military units, seven aircraft types, 35 routes and 30 time periods. The associated linear program has 4,100 rows, 7,400 columns, 39,000 non-zeros, and a maximum row width of three periods. In this scenario, a SWA contingency requires deployment of several Army and Marine Corps brigades from the continental US (CONUS), 15 Air Force fighter wings from CONUS and Europe, and an Army mechanized division from Europe. The movement requirements intentionally exceed delivery capacity in order to strain the system and identify airlift bottlenecks.

Table 1, and Figures 5 and 6 illustrate that solution quality improves with increased

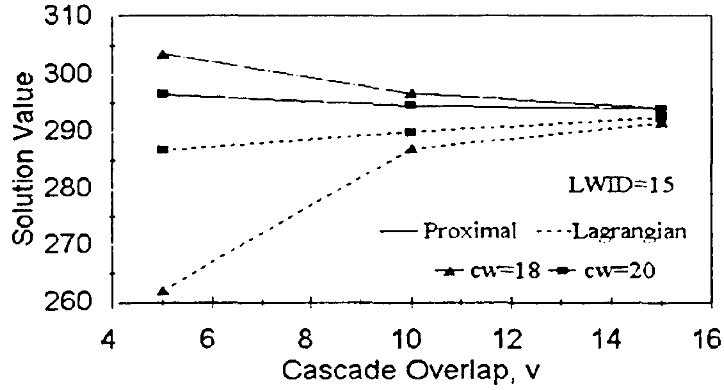

Figure 5: Solution gaps for the Southwest Asia scenario decrease significantly with increased proximal cascade overlap. The triangles show the proximal (solid line) and Lagrangian (dotted line) cascade solution values for an 18-period proximal cascade width; the squares show the solution values for a 20-period width. All Lagrangian cascade widths are 15. The absolute gap, measured by the vertical distance between proximal and Lagrangian solution values, is much smaller with a 10-period overlap than a five-period overlap, and smaller still for a 15-period overlap.

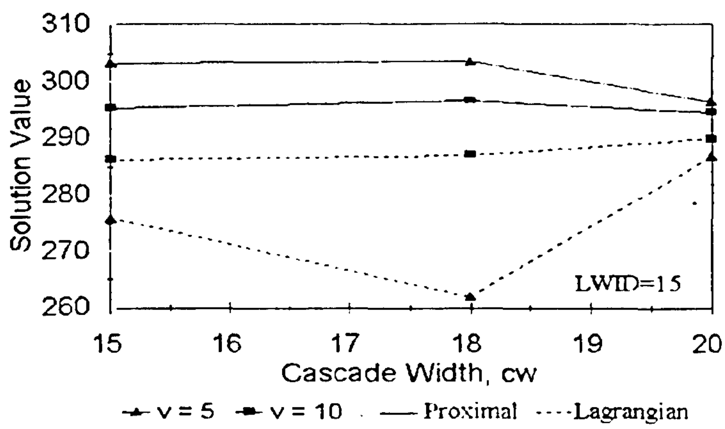

Figure 6: Solution gaps for the Southwest Asia scenario generally decrease as proximal cascade width increases. Proximal cascade width has a smaller effect on the absolute gap than the proximal cascade overlap.

Cascade Width cw	Cascade Overlap v	Upper Bound	Lower Bound	%Gap	Proximal Time (sec)	Lagrange Time (sec)	Total Time (sec)
Monolith		**294.1**	**n/a**	**n/a**	**n/a**	**n/a**	**61**
20	5	296.6	286.9	3.4	47	19	66
20	10	294.6	290.0	1.6	57	20	77
20	15	294.1	292.5	0.6	94	18	112
18	5	303.6	262.0	15.9	46	18	64
18	10	296.7	287.1	3.3	67	21	88
18	15	294.1	291.6	0.9	124	18	142
15	5	303.3	275.9	9.9	41	19	60
15	10	295.4	286.3	3.2	73	19	92
15	12	294.7	285.2	3.3	107	19	126
10	5	305.3	273.7	11.6	41	20	61
10	7	300.0	266.4	12.6	58	20	78

Table 1: Relative gaps and solution times for the Southwest Asia scenario vary with cascade parameter selection. The first two columns show proximal cascade widths and overlaps; Lagrangian cascade widths are all 15. The remaining columns show the performance (computing times are in seconds on an IBM RS6000/590 with 512MB RAM). For example, a proximal cascade with width 18 and overlap 10 gives an upper bound solution value of 296.7; the corresponding Lagrangian lower bound is 287.1, resulting in a gap of 3.3%. The proximal and Lagrange solve times are 67 and 21 seconds, respectively, for a total of 88 seconds. The first row of the table gives the monolith's solution value and time, which provides a baseline for the other runs. Each test uses *CPLEX 3.0* [*CPLEX, 1994*] with primal simplex method and steepest edge pricing.

cascade overlap and width. Figure 5 shows a strictly decreasing gap with increasing cascade overlap for cascade widths of 18 and 20. These decreasing gaps come at a computational cost, however, as indicated by the proximal cascade solution times. Figure 6 also shows generally decreasing gaps with increased cascade width, albeit less convincingly.

6.2 European Infrastructure Scenario

Concurrent with this research, a RAND Corporation study for the Office of the Secretary of Defense (OSD) is examining European air bases transited by USAF airlifters. The purpose of this study is to determine which bases have insufficient infrastructure to adequately support a Major Regional Contingency (MRC) in Southwest Asia [Stucker and Melody, 1996]. The problem consists of 220 military units, six aircraft types, 22 routes, and 30 time

periods. Approximately 75% of the scenario's movement requirements originate in CONUS. The corresponding linear program has 27,000 rows, 126,500 columns, 921,500 non-zeros, and a maximum staircase overlap of two periods.

Cascade Width	Cascade Overlap	Upper Bound	Lower Bound	%Gap	Proximal Time (sec)	Lagrange Time (sec)	Total Time (sec)
Monolith		**106.1**	**n/a**	**n/a**	**n/a**	**n/a**	**980**
20	5	108.7	93.8	15.8	1010	590	1600
20	10	106.9	101.8	5.0	1260	704	1964
20	15	106.9	102.9	3.9	1907	663	2570
18	5	107.4	91.8	17.0	933	630	1563
18	10	107.6	98.3	9.5	1352	605	1957
18	15	107.1	100.4	6.7	2652	715	3367
15	5	109.2	84.5	29.3	959	659	1618
15	10	107.6	96.0	12.1	1527	650	2177
15	12	107.5	99.9	7.6	2307	601	2908
10	5	113.3	75.8	49.4	1061	639	1700
10	7	110.9	83.7	32.4	1483	770	2253

Table 2: Computational results for the European Infrastructure scenario also show that relative gaps and solution times vary with cascade parameter selection. The solve times are much longer than Southewst Asia scenario solve times due to problem size. All runs use the *CPLEX 3.0* Barrier algorithm [*CPLEX*, 1994]. Lagrangian cascade subproblems have 15 periods each. The first row is the monolith baseline; subsequent rows show performance using various proximal cascade parameters. All times are in seconds. Cascading does not save time compared to direct solution of the monolith in the experiments reported in Tables 1 and 2, because the computer used had sufficient memory to solve the monolith with little or no paging. Table 3 shows the time advantage of cascades when memory is limited compared to problem size.

The results of this scenario (see Table 2, and Figures 7 and 8) are generally consistent with those of the first test. Figure 7 shows a pronounced reduction in gap as cascade overlap increases, while Figure 8 shows a more moderate reduction with increased cascade length. Upper bounds are of better quality than lower bounds, due to the sensitivity of the lower bound to small errors in the Lagrangian penalties. Thus, the proximal cascade results show that the effects of myopia are small, since most of the upper bound solution values are within a few percent of the monolith value.

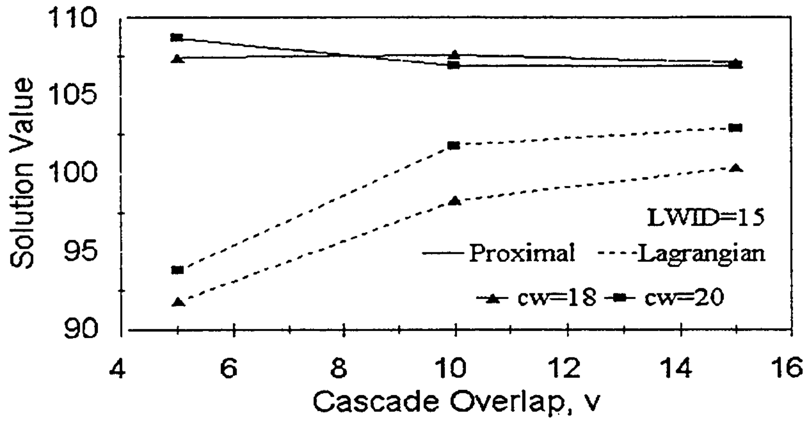

Figure 7: This Figure depicts cascade solution values for the European Infrastructure scenario when proximal cascade overlap is varied. Proximal cascade overlap has as large an effect on this scenario as it did on the notional Southwest Asia scenario. As before, increasing the overlap reduces the gap.

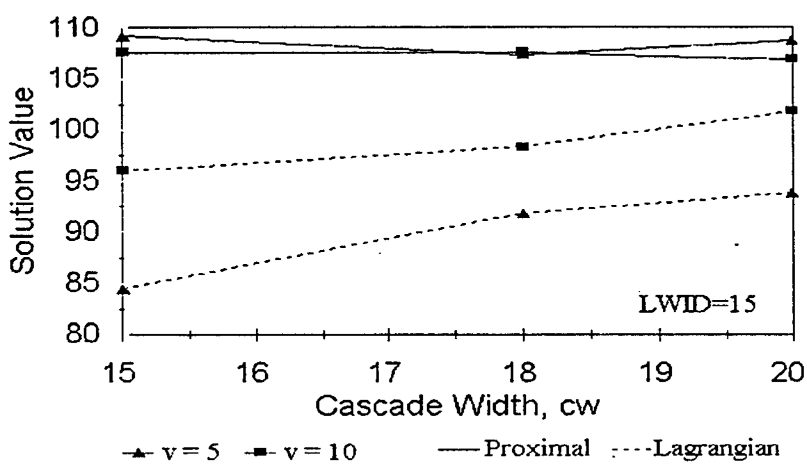

Figure 8: Solution gaps for the European Infrastructure scenario are reduced with increasing proximal cascade width. These reductions, although smaller than those seen in the Southwest Asia scenario, are still quite evident.

6.3 Solve Times

Cascades do not save computing time on the two scenarios described. However, the test platform is a computer with sufficient memory for monolith solution without paging. In order to verify that cascades save time when memory is limited, we constructed two smaller European scenarios by reducing the number of aircraft types, and limiting the number of days that each unit can be delivered. This reduction allows solution by a Dell Pentium Pro 200 mHz desktop computer with 64 MB RAM.

Table 3 shows that cascades save up to 80% of the time required for monolith solution. The savings come at a moderate cost in solution quality, since limited memory requires that cascade subproblems have small widths. This consequence is minor in models such as NRMO, where myopia exists due to wartime uncertainty.

Cascade Width	Cascade Overlap	%Gap	Proximal Seconds	Lagrange Seconds	Total Seconds	% Time Savings
Reduced European Infrastructure I (14,442 rows, 64,252 columns, 462,645 non-zeros):						
Monolith		**n/a**	**n/a**	**n/a**	**4410**	**n/a**
10	5	29.4	572	310	882	80.0
10	7	25.2	844	310	1154	73.8
15	5	14.1	4080	310	4390	0.5
Reduced European Infrastructure II (16,874 rows, 63,336 columns, 453,663 non-zeros):						
Monolith		**n/a**	**n/a**	**n/a**	**4169**	**n/a**
10	5	37.7	532	476	1008	75.8
10	7	19.4	760	480	1240	70.3
15	5	11.4	2160	480	2640	36.7

Table 3: Cascades offer a significant time savings when the monolith cannot be solved with installed memory. The computer used for these results is a Pentium Pro 200 MHz desktop with 64 MB RAM (previous results use an IBM RS6000/590 with 512 MB RAM). The first row of each scenario shows the monolith solution value and time using the *CPLEX* interactive barrier solver [*CPLEX*, 1994]. The next two rows in each scenario indicate cascades offer a dramatic time savings when moderate cascade widths are used. The final row of each scenario shows that much or all of this savings is lost when cascades also require paging.

7 Conclusion

Cascades provide a useful approximation strategy when problem structure permits, and when model size or system myopia warrants. In this paper, we have formalized a cascade method for approximating staircase LPs, and developed a Lagrangian cascade bound for

that approximation. We have also developed the nested-cascade, which may be useful when primal feasibility is difficult to maintain in a traditional nested decomposition.

Using the NRMO model, upper bounds from the proximal cascade are typically within a few percent of monolith optimal. Lower bounds from the Lagrangian cascade have generally less quality, but are often still within a few percent of monolith optimal. Cascade solution times are less than the monolith solution times when small cascade overlaps are used, or when installed memory is limited [Baker, 1997].

8 Acknowledgments

We would like to thank David Morton of the Graduate Program in OperationsResearch at the University of Texas at Austin for his enormous help on both theoretical and applied issues. We are also indebted to Laura Melody Williams, formerly of RAND, now with the Naval Postgraduate School, for her modeling, programming and scenario development skills. As do all adventurers in optimization at the Naval Postgraduate School, we have greatly benefited from the guidance and encouragement of Jerry Brown, Rob Dell, and Kevin Wood.

9 References

Aronson, J., Morton, T., and Thompson, G., "A Forward Simplex Method for Staircase Linear Programs," *Management Science* 31 (1985), 664-679.

Baker, S., *A Cascade Approach for Staircase Linear Programs with an Application to Air Force Mobility Optimization*, Ph.D. Dissertation, Operations Research Department, Naval Postgraduate School, CA, 1997.

Benders, J., "Partitioning Procedures for Solving Mixed Variables Programming Problems," *Numerische Mathematik* 4 (1962), 238-252.

Brooke, A., Kendrick, D., and Meeraus, A., *GAMS, a User's Guide*, The Scientific Press, South San Francisco, 1992.

Brown, G., Dell, R., and Wood, R., "Optimization and Persistence," *Interfaces* 33:5 (Sep-Oct 1997), 15-37.

Brown, G., Graves, G., and Ronen, D., "Scheduling Ocean Transportation of Crude Oil," *Management Science* 33 (1987), 335-346.

Charnes, A., and Cooper, W., *Management Models and Industrial Applications of Linear Programming*, John Wiley & Sons, New York, 1961.

CPLEX Optimization Inc., *CPLEX,* version 3.0 callable library, Incline Village NV, 1994.

Damm, E., *Comparison of the Force Deployment Estimator and the NPS/RAND Mobility Optimizer*, Masters Thesis, Operations Research Department, Naval Postgraduate School, CA, forthcoming September 1998.

Department of the Air Force, *Basic Aerospace Doctrine of the United States Air Force*, Air Force Manual 1-1, Vol. II, US Government Printing Office, 1992.

Glassey, C., "Nested Decomposition and Multi-Stage Linear Programs," *Management Science* 20 (1973), 282-292.

Ho, J., and Manne, A., "Nested Decomposition for Dynamic Models," *Mathematical Programming* 6 (1974), 121-140.

Killingsworth, P., and Melody, L., "CONOP Air Mobility Optimization Model," internal memorandum, RAND Corp., Santa Monica CA, July 1994.

Manne, A., "Sufficient Conditions for Optimality in an Infinite Horizon Development Plan," *Econometrica* 38 (1970), 18-38.

Melody, L., Baker, S., Rosenthal, R., Morton, D., and Killingsworth, P., *NPS/RAND Mobility Optimization (NRMO): A New Model for Air Mobility Analysis*, Project Memorandum PM-625-AF, RAND Corp., Santa Monica CA, May 1997.

Morton, D., Rosenthal, R., and Lim Teo Weng, "Optimization Modeling for Airlift Mobility," *Military Operations Research*, 1, No. 4, Winter 1996.

Parker, R., and Rardin, R., *Discrete Optimization*, Academic Press, Boston, 1988.

Rosenthal, R., Baker, S., Lim, T., Fuller, D., Goggins, D., Toy, A., Turker, Y., Horton, D., Briand, D., and Morton, D., "Application and Extension of the THRUPUT II Optimization Model for Airlift Mobility," *Military Operations Research*, 3, No. 2, 1997.

Rosenthal, R., Morton, D., Melody., L, and Baker, S., "NPS/RAND Mobility Optimizer," internal memorandum, Naval Postgraduate School, Monterey, CA, 1997.

Stucker, J., and Melody, L., "Impacts of European Airbase Infrastructure on Airlift Mobility," RAND study for the Office of the Secretary of Defense, 1997.

Walker, S., *Evaluating End Effects for Linear and Integer Programs Using Infinite-Horizon Linear Programming*, Ph.D. Dissertation, Operations Research Department, Naval Postgraduate School, CA, 1995.

DISTRIBUTION LIST

1. Research Office (Code 09) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Dudley Knox Library (Code 013) 2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Defense Technical Information Center 2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

4. Therese Bilodeau (Editorial Assistant) 1
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Prof. Richard E. Rosenthal (Code OR/Rl) 5
   Chairman, Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Lt Col Steven F. Baker 5
   HQ USAFA/DFM
   2354 Fairchild Drive, Suite 6H94
   USAF Academy, CO 80840

7. Chief, Global Mobility Division 4
   Air Force Studies & Analyses Agency
   1570 Air Force, Pentagon
   Washington, DC 20330

8. Dr. Neal Glassman 1
   Air Force Office of Scientific Research
   (Code NM)
   110 Duncan Avenue, Suite B115
   Bolling AFB, DC 20332-8080

9. Dr. Donald Wagner 1
   Office of Naval Research
   (Code 1111)
   800 North Quincy Street
   Arlington, VA 22217